ショートメッセージの部分空間表現と宛先推定

清水 彰人[1]　酒井 智弥[2]　小林 透[2]

概要：ソーシャルネットワーキングサービス（SNS）の多くは，メッセージを交換する際に，その宛先や話題（ハッシュタグ等）を明示しながら使うことが前提となっている．端末の操作に不慣れな利用者や高齢者からのメッセージの配信を支援するためには，宛先が明示されていない短いメッセージの内容から宛先を推定する機能を実現するべきである．しかし，メッセージの例が大量に入手できなければ，単語の出現頻度やベイズ則に基づく既存の文書分類の技術を応用できない．この課題に対して，本稿では，メッセージを構成している単語の組合せが表す意味の類似性に基づく宛先推定法を提案する．提案手法は，単語の意味を線形代数的に演算可能にするベクトル表現 word2vec を利用し，短いメッセージと宛先候補をそれぞれ線形ベクトル部分空間で表すことで，相互部分空間法による類似性の定量化を実現する．この手法により，わずか数語の短いメッセージから宛先が推定可能であることを示唆する単純な実験例を示す．

Recognizing Destinations of Short Messages via Subspace Representation

AKIHITO SHIMIZU[1]　TOMOYA SAKAI[2]　TORU KOBAYASHI[2]

1. はじめに

Twitter，Facebook，LINE をはじめとする SNS が普及している．高齢者や，端末の操作に不慣れな利用者が容易に SNS を利用できるようになれば，若年者をはじめとする他者とメッセージを交換することによる双方向のコミュニケーションが促進される．SNS の多くは，メッセージの宛先としてユーザやグループを指定したり，ハッシュタグ等で話題のメタ情報を明示しながら，主にテキストを流通させることが前提となっている．ゆえに，例えば高齢者とのコミュニケーションに SNS を応用するためには，音声認識等で高齢者からのメッセージをテキスト化するだけでなく，メッセージの宛先指定や話題の明示を支援する機能の付加も必要となる．

高齢者からのメッセージの宛先推定は，迷惑メールフィルタや文書分類で成功している機械学習の既存手法を流用して解決できる課題ではない．例えば，単純ベイズ分類器（naive Bayes classifier）や tf-idf のように，既存手法はメッセージ中の単語の出現頻度や重要度を見積もるために大量のメッセージの例を必要とする．ところが，高齢者が SNS を活用していない現状で，高齢者からのメッセージの例を大量に用意することは非現実的である．また，高齢者からのメッセージは，主に少数の単語から構成される短文であると思われる．ゆえに，宛先に関係する単語が予め既知であったとしても，その単語がメッセージ中で必ずしも使われているとは限らない．

ショートメッセージの宛先を推定する課題を解決する糸口は，出現頻度等で定量化した個々の単語の重要性ではなく，単語の組合せが表す意味に着目することにあると考えられる．例えば，高齢者から親族宛のメッセージに関連しそうなキーワードとして “正月”，“帰省”，“孫” 等を想定していれば，高齢者からのメッセージ「今度，新幹線で帰っておいで」が親族宛であることは容易に推定できる．キーワードがメッセージに出現しなくても，“新幹線” と “帰っ” の組合せは “帰省” の意味から遠くない．このように単語やその組合せが表す意味の類似性を計量できれば，ショートメッセージから宛先を推定する手法を設計できるであろう．

本稿では，word2vec [1,2] と呼ばれる単語のベクトル表現を利用したショートメッセージの宛先推定法を提案する．単

[1] 長崎大学工学部
[2] 長崎大学大学院工学研究科

語のベクトル表現 word2vec は，単語の意味を線形代数的に演算可能にする技術である．単語 w を表すベクトルを $\boldsymbol{v}(w)$ とすると，例えば $\boldsymbol{v}(\text{“king”}) - \boldsymbol{v}(\text{“man”}) + \boldsymbol{v}(\text{“woman”})$ の演算で $\boldsymbol{v}(\text{“queen”})$ の近似ベクトルが作られるような性質がある．このような単語のベクトル表現を利用すると，少数の単語の組合せと見なせるショートメッセージは，それらの単語ベクトルが表す線形ベクトル部分空間として表せる．ショートメッセージの潜在的意味はこの部分空間に対応づけられ，意味が近いショートメッセージの部分空間は互いに類似していると考えられる．宛先についても同様に，少数のキーワードが与えられれば線形ベクトル部分空間で表現できる．本研究の新規性は，単語の意味の類似性を単語ベクトルで計量するのではなく，単語の組合せが表す意味の類似性を部分空間を用いて定量化する点にある．

単語のベクトル表現 word2vec およびショートメッセージと宛先の部分空間表現について 2 節で詳細を解説する．また，3 節では，部分空間どうしの類似度を計算する方法および最大類似度に基づく宛先推定を実現する相互部分空間法 [3] について解説する．この手法により，わずか数語のショートメッセージから宛先が推定可能であることを示唆する単純な実験例を 4 節に示す．

2. 単語の組合せの部分空間表現

2.1 単語のベクトル表現

Mikolov らが提案した word2vec [1,2] は，「似た文脈に出現する単語は互いに似た意味を持つ」という分布仮説に基づき，文脈から単語を予測する CBOW（continuous bag-of-words）モデルまたは単語から文脈を予測する skip-gram モデルによって単語の分散表現 [4] を効率的に教師なし学習する手法である．Skip-gram モデルでは，大量の文書から任意に抜き出した連続する単語の列について，ある単語とその周囲の単語が類似した実ベクトルで表されるように設計されている．

文書中の単語 $c \in \mathrm{V}$（V は語彙）とその前後 L 個以内にある単語 $w \in \mathrm{V}$ の組 (w, c) の多重集合を D とする．また，D に含まれる組 (w, c) の数を $\#(w, c)$，単語 c の出現頻度を $P_{\mathrm{D}}(c) = \#(c)/|\mathrm{D}| = \sum_{w \in \mathrm{V}} \#(w, c) / \sum_{(w,c) \in \mathrm{D}} 1$ とする．Skip-gram モデルに基づき単語 w を表す d 次元ベクトル $\boldsymbol{v}(w) \in \mathbb{R}^d$ を求める問題は，次式の最大化問題として表せる [5].

$$\max_{\boldsymbol{v}(w), \forall w \in \mathrm{V}} \sum_{(w,c) \in \mathrm{D}} \#(w, c) (\log \sigma(\boldsymbol{v}(w)^\top \boldsymbol{v}(c)) + k \mathrm{E}_{c_N \sim P_{\mathrm{D}}} [\log \sigma(-\boldsymbol{v}(w)^\top \boldsymbol{v}(c_N))]) \quad (1)$$

ただし，$\sigma(x) = 1/(1 + e^{-x})$ はシグモイド関数，E は期待値を表す．式 (1) の第 1 項は，文脈に共起する単語 c と w の類似度（単語ベクトルの内積）が大きく正になることを奨励している．この項のみでは単語ベクトルがすべて同一の向きを持つ自明な解を持つので，第 2 項は，文脈に関係なく出現頻度 P_{D} に従ってランダムに選ばれた単語 c_N との類似度がなるべく大きく負になる（単語ベクトルの逆向きになる）ように正則化している．これは負例サンプリング（negative sampling）と呼ばれる．k は負例の数であり，正則化の重みに相当する．

なお，word2vec では，実装において雑多な技法が導入されている．例えば，窓サイズ L は，単語列の実際の長さではなく，最大の長さのことであり，それ以下で任意の長さが実際に使われている．また，極端に出現頻度が低い・高い単語を扱わないための閾値指定がある．頻度の高い単語（例えば，英文における a, the 等の冠詞）を無視すれば，語彙が減る一方で，実質的に単語列の長さが伸びる効果がある．

2.2 メッセージと宛先の部分空間表現

ひとつのメッセージは複数の単語から構成されている．単語の意味を線形合成できる word2vec によって単語をベクトル表現すると，メッセージを構成する n 個の単語ベクトルの多重集合 $\mathrm{S} = \{\boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(n)} \in \mathbb{R}^d\}$ によって実空間 $\mathbb{R}^d$ の線形ベクトル部分空間

$$\mathcal{S} = \operatorname{span} \mathrm{S} = \{\boldsymbol{v} | \ \boldsymbol{v} = \sum_{j=1}^{n} a_j \boldsymbol{v}^{(j)}, \ \forall a_j \in \mathbb{R}, \ \forall \boldsymbol{v}^{(j)} \in \mathrm{S}\} \quad (2)$$

を定義できる．部分空間 $\mathcal{S}$ は，メッセージに現れる単語のベクトルからなる線形結合全体をなす集合であり，単語の組合せが表すメッセージの意味を表現しているものと見なせる．同様に，宛先を記述するキーワードが何らかの方法で得られれば，キーワードの組合せが表す意味を部分空間で表現できる．

単語ベクトルが張る線形ベクトル部分空間の正規直交基底を求める基本的な算法は特異値分解（singular value decomposition; SVD）である [6]．メッセージを構成する n 個の単語ベクトルを列に並べた行列をデータ行列

$$\boldsymbol{X} = \left[\boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(n)}\right] \in \mathbb{R}^{d \times n}$$

と定義する．ランク $r = \operatorname{rank} \boldsymbol{X} \leq \min(d, n)$ のデータ行列 $\boldsymbol{X}$ の薄い特異値分解（thin SVD）は，

$$\boldsymbol{X} = \boldsymbol{U} \operatorname{diag}(\boldsymbol{\kappa}) \boldsymbol{V}^\top$$

である．ただし，

$$\boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{V}^\top \boldsymbol{V} = \boldsymbol{I}_r$$

であり，$\boldsymbol{I}_r$ は r 次の単位行列，$\operatorname{diag}(\boldsymbol{\kappa})$ は特異値 $\boldsymbol{\kappa} = [\kappa_1, \ldots, \kappa_r]^\top \in \mathbb{R}_+^r$ からなる対角行列を表す．左特異ベクトルを列に持つ行列 $\boldsymbol{U} \in \mathbb{R}^{d \times r}$ は線形ベクトル部分空間 $\mathcal{S}$ の基底行列であり，$\mathcal{S}$ を張る r 本の d 次元正規直交基底ベクトルを与える．

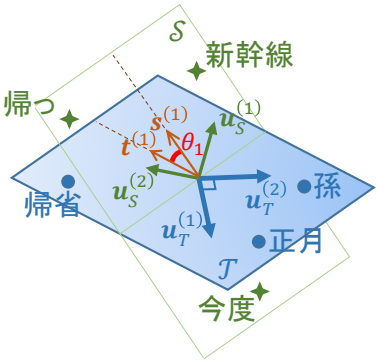

図 1　メッセージ「今度，新幹線で帰っておいで」の部分空間 $\mathcal{S}$ と，キーワード（“正月”，“帰省”，“孫”，…）を設定した宛先の部分空間 $\mathcal{T}$ の間の類似度評価．$\mathcal{S}$ と $\mathcal{T}$ の基底ベクトルはそれぞれ $\boldsymbol{U}_S = \left[\boldsymbol{u}_S^{(1)}, \boldsymbol{u}_S^{(2)}\right]$ $\boldsymbol{U}_T = \left[\boldsymbol{u}_T^{(1)}, \boldsymbol{u}_T^{(2)}\right]$ である．$\mathcal{S}$ と $\mathcal{T}$ の第 1 正準ベクトル $\boldsymbol{s}^{(1)}$ と $\boldsymbol{t}^{(1)}$ のなす第 1 正準角 θ_1 が小さいほど類似度が高い．

3.　相互部分空間法による宛先推定

メッセージの部分空間と宛先の部分空間の間の類似度を計算できれば，メッセージに対して最も類似度が高い宛先を推定できる．正規直交の基底行列 $\boldsymbol{U}_S \in \mathbb{R}^{d \times r_S}$ と $\boldsymbol{U}_T \in \mathbb{R}^{d \times r_T}$ で張られる部分空間 $\mathcal{S}$ と $\mathcal{T}$ の間の類似度は，正準角 θ_k $(0 \leq \theta_1 \leq \cdots \leq \theta_m, m = \min(r_S, r_T))$ によって定量化される [3,7–9]．正準角の余弦は次式で定義される．

$$\cos\theta_k = \max_{\boldsymbol{s}^{(k)} \in \mathcal{S}} \max_{\boldsymbol{t}^{(k)} \in \mathcal{T}} {\boldsymbol{s}^{(k)}}^\top \boldsymbol{t}^{(k)} \tag{3}$$

$$\text{subject to } \left\|\boldsymbol{s}^{(k)}\right\|_2 = \left\|\boldsymbol{t}^{(k)}\right\|_2 = 1,$$

$${\boldsymbol{s}^{(k)}}^\top \boldsymbol{s}^{(l)} = {\boldsymbol{t}^{(k)}}^\top \boldsymbol{t}^{(l)} = 0 \, (l = 1, \ldots, k-1) \tag{4}$$

最大値を与える単位ベクトル $\boldsymbol{s}^{(k)} \in \mathcal{S}$ と $\boldsymbol{t}^{(k)} \in \mathcal{T}$ はそれぞれ $\mathcal{S}$ と $\mathcal{T}$ の正準ベクトルと呼ばれる．特に，正準ベクトル $\boldsymbol{s}^{(1)}$ と $\boldsymbol{t}^{(1)}$ のなす第 1 正準角 θ_1 は，図 1 のように部分空間 $\mathcal{S}$ と $\mathcal{T}$ のなす最小角である．メッセージと宛先が共通の単語を持たなくても，両者の意味が近ければこの正準角が小さいことが期待できる．相互部分空間法 [3] では，第 1 正準角の余弦 $\cos\theta_1$ が類似度として用いられる．正準角の余弦 $\cos\theta_k$ は，基底行列の積 $\boldsymbol{U}_S^\top \boldsymbol{U}_T \in \mathbb{R}^{r_S \times r_T}$ の特異値として計算できる [7]（付録参照）．

与えられたメッセージに対する宛先を相互部分空間法によって推定する提案手法は下記の手順である．

(1) 大量の文書から word2vec で単語ベクトルを作成しておく．メッセージや宛先に関係なく，一般的な単語の意味の関係を獲得できればよい．

(2) C 種類の宛先 T_j $(j = 1, \ldots, C)$ それぞれについて，事前に集めたキーワードの単語ベクトルを列に持つデータ行列 $\boldsymbol{X}_{T_j}$ を特異値分解し，あらかじめ各宛先の基底行列 $\boldsymbol{U}_{T_j}$ を求めておく．

(3) メッセージ S が与えられたら，その単語ベクトルを列に持つデータ行列 $\boldsymbol{X}_S$ を特異値分解し，メッセージの基底行列 $\boldsymbol{U}_S$ を計算する．

(4) 類似度が最大となる宛先 T_{pred} を探す．

$$T_{\text{pred}} = \arg \max_{T \in \{T_1, \ldots, T_C\}} \sigma_1(\boldsymbol{U}_S^\top \boldsymbol{U}_T)$$

$\sigma_1(\boldsymbol{U}_S^\top \boldsymbol{U}_T)$ は行列 $\boldsymbol{U}_S^\top \boldsymbol{U}_T$ の最大特異値を表す．

4.　宛先推定の実験例

宛先としてユーザグループをいくつか仮定し，それぞれの宛先に数語のキーワードを設定したとき，メッセージとその適切な宛先の間の類似度が最も高くなることを確認する単純な実験例を示す．

(1) 2015 年 11 月 2 日時点の日本語 Wikipedia の全記事 [*1] の平文を，形態素解析器 MeCab（辞書は NAIST-jdic）で分かち書きし，word2vec (Revision 42)[*2] の訓練データとした．単語ベクトルの次元数は $d = 200$ とし，他のパラメタは既定値である．結果，405,565 語の単語ベクトルが得られた．

(2) 宛先 T_j として，「病院」，「宅配」，「親族」のユーザグループを仮定する（$C = 3$）．著者が任意に列挙した各宛先のキーワードを表 1 に示す．

表 1　宛先とキーワード．

病院 T_1	医者, 病院, 薬, 処方箋, ケア, 通院, 患者, 救急車, 看護, 手術, 安静, 薬学, クリニック, 医療, 福祉, 介護, 内科, 診察, 白衣, 院長, ベッド, 通院
宅配 T_2	宅配, コンビニ, 出前, 寿司, 配達, 電話, ラーメン, 餃子, 中華, 料理, 野菜, 値段
親族 T_3	帰省, 母, 娘, 盆, 夏休み, 正月, 家族, 父, 祖父, 祖母, 渋滞, 孫, 墓, 跡継ぎ, 実家, 田舎, 息子

(3) 各宛先を想定して作成したショートメッセージ S_i を表 2 に示す．分かち書きされているすべての単語から各メッセージの部分空間を作成した．

表 2　各宛先へのメッセージ．

病院宛 S_1	先生 , 子供 が 熱 出し た ので 診 て ください.
宅配宛 S_2	弁当 の 注文 を おねがい し ます.
親族宛 S_3	今度 , 新幹線 で 帰っ て おい で.

(4) 各メッセージと宛先の間の類似度 $\sigma_1(\boldsymbol{U}_{S_i}^\top \boldsymbol{U}_{T_j})$ を表 3 に示す．宛先のキーワードと一致する単語がメッセージに含まれていないにもかかわらず，各メッセージは適切な宛先に対して類似度が最も大きくなっている．

病院宛メッセージ S_1 は，親族 T_3 に対しても類似度が高い．その原因として，メッセージ S_1 に含まれる単語 “子供” と意味が近い “息子” 等の単語が親族 T_3 のキーワードに挙げられていることが考えられる．宅配宛メッセージ

[*1] https://dumps.wikimedia.org/jawiki/20151102/
[*2] http://word2vec.googlecode.com/

表 3　メッセージと宛先の間の類似度.

メッセージ＼宛先	病院 T_1	宅配 T_2	親族 T_3
病院宛 S_1	**0.84**	0.55	0.77
宅配宛 S_2	0.49	**0.85**	0.57
親族宛 S_3	0.52	0.59	**0.65**

S_2 は，宅配 T_2 に対して明確に高い類似度を示している．メッセージ S_2 に含まれる単語 “弁当” に意味が近い食べ物に関連するキーワードが宅配 T_2 のみに挙げられているためと考えられる．親族宛メッセージ S_3 について，親族 T_3 は他の宛先と比べて類似度が顕著に高くない．メッセージ S_3 に含まれる単語 “帰っ” は親族 T_3 のキーワード “帰省” と意味が近いが，“新幹線” という単語が病院 T_1 や宅配 T_2 に挙げられている乗り物や移動に関連するキーワード “救急車”，“配達”，“出前” に意味が近いからであろう．

5. おわりに

本稿では，単語の組合せが表す意味の類似性を計量することで，わずか数語の短いメッセージから宛先を推定可能にする手法を提案した．提案手法では，単語の意味を線形代数的に演算可能にする word2vec を利用してメッセージと宛先を部分空間で表現している．部分空間を用いて単語の組合せが表す意味の類似性を評価する本手法は，単語のベクトル表現 word2vec で個々の単語の間の類似性を評価する先行研究よりも合理的である．また，メッセージと宛先の対応関係を機械学習するための訓練データは不要なので，SNS を活用していない高齢者等を対象としたメッセージ交換への応用に適している．

実験例では宛先のキーワードを著者が用意したが，実用上は事前に交換されたメッセージの例から抽出するべきであろう．または，宛先推定の都度，そのメッセージからキーワードを抽出し，宛先部分空間の基底を更新する算法も設計可能である．

実際のショートメッセージについて宛先推定の性能を調整または検証するためには，統計的な評価に十分な数のメッセージ交換の実例を用意する必要がある．また，word2vec で単語ベクトルを作成する際に，分かち書きした単語の正規化や，word2vec のパラメタ調整も検討の余地がある．

謝辞　本研究は総務省 SCOPE (9254) の助成を受けている．

参考文献

[1] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” arXiv preprint arXiv:1301.3781, 2013.
[2] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality,” Advances in Neural Information Processing Systems, pp.3111–3119, 2013.
[3] 前田賢一，渡辺貞一，“局所的構造を導入したパターン・マッチング法，” 電子情報通信学会論文誌 D，vol.68，no.3，pp.345–352，1985.
[4] G.E. Hinton, “Learning distributed representations of concepts,” Proceedings of the 8th Annual Conference of the Cognitive Science Society, pp.1–12, 1986.
[5] O. Levy and Y. Goldberg, “Neural word embedding as implicit matrix factorization,” Advances in Neural Information Processing Systems, pp.2177–2185, 2014.
[6] G.H. Golub and C.F. Van Loan, Matrix Computations (3rd Ed.), Johns Hopkins University Press, Baltimore, MD, USA, 1996.
[7] k. Björck and G.H. Golub, “Numerical methods for computing angles between linear subspaces,” Mathematics of Computation, vol.27, no.123, pp.579–594, 1973.
[8] A. Edelman, T.A. Arias, and S.T. Smith, “The geometry of algorithms with orthogonality constraints,” SIAM Journal on Matrix Analysis and Applications, vol.20, no.2, pp.303–353, 1998.
[9] J. Hamm and D.D. Lee, “Grassmann discriminant analysis: a unifying view on subspace-based learning,” Proceedings of the 25th International Conference on Machine LearningACM, pp.376–383 2008.

付　　録

部分空間 $\mathcal{S}$，$\mathcal{T}$ それぞれの正規直交基底 $\boldsymbol{U}_S$，$\boldsymbol{U}_T$ を用いて，正準ベクトルをそれぞれ

$$\boldsymbol{s}^{(k)} = \boldsymbol{U}_S\boldsymbol{\phi}^{(k)} \in \mathcal{S}, \qquad \boldsymbol{t}^{(k)} = \boldsymbol{U}_T\boldsymbol{\psi}^{(k)} \in \mathcal{T}$$

と書き表せる．式 (4) より，これらの正準ベクトルはそれぞれ互いに直交する単位ベクトルであるから，$\boldsymbol{\phi}^{(k)} \in \mathbb{R}^{r_S}$，$\boldsymbol{\psi}^{(k)} \in \mathbb{R}^{r_T}$ もそれぞれ次式のように互いに直交する単位ベクトルである．

$$\left\|\boldsymbol{\phi}^{(k)}\right\|_2 = \left\|\boldsymbol{\psi}^{(k)}\right\|_2 = 1 \quad (k = 1, \ldots, m),$$
$$\boldsymbol{\phi}^{(k)\top}\boldsymbol{\phi}^{(l)} = \boldsymbol{\psi}^{(k)\top}\boldsymbol{\psi}^{(l)} = 0 \quad (l = 1, \ldots, k-1)$$

ゆえに，式 (3) は以下のように書き表せる．

$$\begin{aligned}\cos\theta_k &= \max_{\boldsymbol{u}^{(k)}\in\mathcal{S}} \max_{\boldsymbol{v}^{(k)}\in\mathcal{T}} \boldsymbol{u}^{(k)\top}\boldsymbol{v}^{(k)} \\ &= \max_{\boldsymbol{\phi}^{(k)}\in\mathbb{R}^{r_S}} \max_{\boldsymbol{\psi}^{(k)}\in\mathbb{R}^{r_T}} (\boldsymbol{U}_S\boldsymbol{\phi}^{(k)})^\top\boldsymbol{U}_T\boldsymbol{\psi}^{(k)} \\ &= \max_{\boldsymbol{\phi}^{(k)}\in\mathbb{R}^{r_S}} \max_{\boldsymbol{\psi}^{(k)}\in\mathbb{R}^{r_T}} \boldsymbol{\phi}^{(k)\top}\boldsymbol{U}_S^\top\boldsymbol{U}_T\boldsymbol{\psi}^{(k)}\end{aligned}$$

ここで，$\boldsymbol{U}_S^\top\boldsymbol{U}_T$ の特異値分解を

$$\begin{aligned}\boldsymbol{U}_S^\top\boldsymbol{U}_T &= \boldsymbol{P}\,\mathrm{diag}(\boldsymbol{\sigma})\boldsymbol{Q}^\top \\ \boldsymbol{\sigma} &= [\sigma_1, \ldots, \sigma_m]^\top \\ \boldsymbol{P} &= [\boldsymbol{p}^{(1)}, \ldots, \boldsymbol{p}^{(m)}] \in \mathbb{R}^{r_S\times m} \\ \boldsymbol{Q} &= [\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(m)}] \in \mathbb{R}^{r_T\times m}\end{aligned}$$

とすると，正準角の余弦 $\cos\theta_k$ は第 k 特異値と一致する．

$$\begin{aligned}\cos\theta_k = \max_{\boldsymbol{\phi}^{(k)}\in\mathbb{R}^{r_S}} \max_{\boldsymbol{\psi}^{(k)}\in\mathbb{R}^{r_T}} &\boldsymbol{\phi}^{(k)\top}\boldsymbol{P}\,\mathrm{diag}(\boldsymbol{\sigma})\boldsymbol{Q}^\top\boldsymbol{\psi}^{(k)} \\ &\text{subject to } \left\|\boldsymbol{\phi}^{(k)}\right\|_2 = \left\|\boldsymbol{\psi}^{(k)}\right\|_2 = 1, \\ &\boldsymbol{\phi}^{(k)\top}\boldsymbol{\phi}^{(l)} = \boldsymbol{\psi}^{(k)\top}\boldsymbol{\psi}^{(l)} = 0\ (l = 1, \ldots, k-1) \\ = \sigma_k \quad &(\boldsymbol{\phi}^{(k)} = \boldsymbol{p}^{(k)}, \boldsymbol{\psi}^{(k)} = \boldsymbol{q}^{(k)}\text{のとき}) \qquad \square\end{aligned}$$